打造正面宣传效果的量化公器
——广东佛山创新应用媒体大数据推出网络正能量指数

范春柏
（佛山市网络安全和信息化协会，广东 佛山 528000）

摘 要：能否有一个办法把网络中对某一单位的正面宣传报道做出一个较为全面的统计？能否量化网络正能量传播的效果？结合以上需求，运用媒体大数据、人工智能、机器学习，佛山市网络安全和信息化协会创新研发出一套网络正能量指数系统，对全市 270 余个党政机关进行指数测评。对进一步把握网络传播规律，加强网络正能量传播的时、度、效，为市、区、镇开展网络正能量传播工作、共同营造清朗的网络空间提供科学的数据支撑和有力的信息支持。

关键词：网络正能量指数；网络传播规律；媒体大数据；机器学习；权重　　**中图分类号：**G641　　**文献标识码：**A
文章编号：1671-0134（2021）08-029-03　　DOI：10.19483/j.cnki.11-4653/n.2021.08.006
本文著录格式：范春柏 . 打造正面宣传效果的量化公器——广东佛山创新应用媒体大数据推出网络正能量指数 [J]. 中国传媒科技，2021（08）：29-31.

对很多党政机关部门来说，宣传工作都是非常重要的，不仅关系到一个单位的政绩能否得到大众认可，还关系到网上群众路线走得好不好的问题。

很多单位在做了一些正能量活动和举措之后，不清楚哪些媒体发了哪些稿、稿件什么样，经常有统计不全面的情况出现。因为，有些新闻稿件在百度中是不大容易搜到的，况且，目前渠道众多，PC 端只是其中一个出口，更有手机 App、微信、微博、入驻号等等。能否有一个办法把所有出口中对某一单位的正面宣传报道做出一个较为全面的统计呢？

另外，全面贯彻落实习近平新时代中国特色社会主义思想，特别是习近平总书记关于网络强国 [1] 的重要思想，紧扣“网聚正能量”主题，深入推进“网络正能量传播提升工程”，是各级网信办的重要工作。网络正能量传播做得好不好，哪些单位做得好，需要一个可量化的指标。

在这两种需求下，结合媒体大数据应用，在佛山市委网信办指导下，佛山市网络安全和信息化协会创新研发出一套网络正能量指数系统。[2] 对全市市级、区级、镇街的党政机关进行网络正能量指数测评。该指数评价系统在 2019 年 11 月获得了软件著作权证书。

指数系统运行近两年时间，充分发挥了指数的指标功能、指引作用，对进一步把握网络传播规律，进一步加强网络正能量传播的时、度、效，为市、区、镇开展网络正能量传播工作、共同营造清朗的网络空间提供科学的数据支撑和有力的信息支持。

佛山网络正能量指数系统这一创新做法，也得到了广东省委网信办和中央网信办的肯定。

1. 网络正能量报道数据从哪来

首先要明确哪些属于网络正能量内容。在重大政策执行、重大主题、重大活动、重大事件、重要节点、热点问题和突发事件中发挥网络正能量引导作用的相关媒体报道，倡导广大网民自觉传播和弘扬正能量，营造清朗的网络空间，还包括有利于弘扬社会主义核心价值观的报道内容。

简单来说：“社会涉及面较广的事件，通过媒体和网络媒体报道出来的，具有正面促进作用的”都算为正能量内容。但是，本单位或部门的职能工作，包括：参观学习、会议、研讨、政策发布等不计算在内。

采用互联网数据采集技术，把网上最新的相关信息获取到数据库里，再通过语义分析结合关键词库等规则，把所有有关佛山的正能量内容归集起来，为下一步的数据计算做准备。

第一步：通过后台设置，确定定向采集任务。

第二步：针对不同的目标数据源，按照来源的内容标签来进行相应的采集配置，其中分为获取 API 数据接口和利用正则表达式进行页面内容分解两种方式。

第三步：调度采集任务，实现与目标站点同步更新，增量采集。

第四步：采集到数据结果后，完成数据异构到同构的过程。

第五步：把采集到的内容进入下一步分析处理。

2. 指数值是如何计算出来的？

相关内容获取之后，需要对内容进行分析处理，从抓取的信息里面挑选出有关佛山及正能量的内容，分析层主要包含两个方面，一是页面元素分析，二是相关性分析。

2.1　页面元素分析

在页面采集到以后，要从中提取出链接、元数据、正文、标题、摘要来，以便进行后续的过滤和其他处理。以标题为例，其判断方法如下图 1。

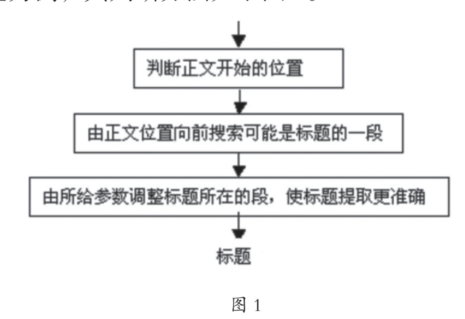


图 1

2.2 相关性分析

首先将所采集内容进行分词处理，提取出现频率较高的关键词，进行入库预处理（为后面的“关键词学习”部分进行词汇储备）。分词完成后，内容关键词与预设关键词进比对，若符合规则，则为目标内容，即有关佛山正能量的内容，最后进行入库处理。

在预设关键词的规则模型里，可设置“与”“非”“或”三种组合状态，即必须某两个或多个关键词同时出现，例如地标（佛山）+ 部门（网信办或文明办或……），“非”即不包含某个关键词。

网络正能量指数系统分为抓取、分析、呈现，三个部分，公众可以看到呈现端，也就是各单位网络正能量指数的排行和具体数值。排名靠前的单位还会有简明的数据分析，以此为该单位和其他单位下一步正能量传播工作做好指引（图 2）。

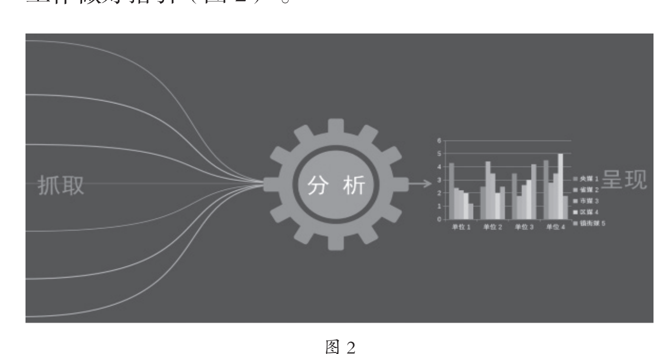


图 2

举个例子（图 3），当数据池里有了大量的数据之后，

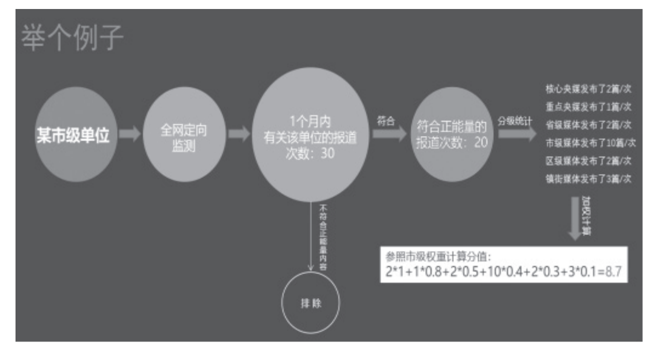


图 3

就可以利用这些数据进行标准量化，这就是我们指数的算法模型。里面说的“报道次数”，是以文章出现的“节点数”来算，如 1 篇文章在 A、B、C 三个媒体发布了，其中 A 媒体在不同栏目里共推荐了 3 次，B、C 各 1 次，那么这个报道次数就为 5 次。

要说明网络正能量指数的算法，先定义一个概念：权重。

2.2.1 权重是什么？

权重是计算指数的重要指标，根据各个部门级别的不同，设定相应权重，尽可能达到更公平、公正的评分准则。

根据全市 5 区和各镇街单位的行政级别，对应不同等级媒体设定不同的权重。数据源分为核心央媒、重点央媒、省媒、市媒、区媒及镇（街）媒体 6 个级别，数据源分别包括网站、App、微信公众号、微信等。

数据源分级如下：

核心央媒：

人民网、新华网、央视网、央广网等。

重点央媒：

光明网、中青网、环球网、中国网、中新网等。

省级媒体：

南方网、南都（奥一）、南方 +（首页）、澎湃、界面、北京时间、封面等。

市级媒体：

佛山日报（官网）、电视台（官网）、电台（官网）、今日头条省级频道、南方 +（频道、地方推荐）等。

区级媒体：

区电视台、今日头条（城市、入驻号）、门户（新浪、腾讯、网易、搜狐）城市频道转载、入驻号等。

镇街媒体：

各社区报、镇（街）官方网站和公众号等。

权重等级设定由大到小顺序为：核心央媒 > 重点央媒 > 省级媒体 > 市级媒体 > 区级媒体 > 镇街媒体，具体参看图 4。

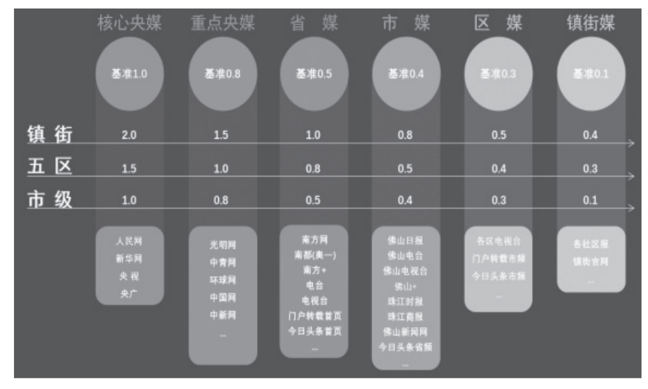


图 4

而对应各级单位，各级媒体的权重也不同。所以，指数系统为市、区、镇（街）分别设置了不同的权重（在

同等条件下，镇街和区级单位要获得央媒报道相对市级单位来说会更难一些，所以权重也需相应提高）。

2.2.2　*在哪些媒体发布才算数*

媒体大数据[3]的信源包括：核心央媒、重点央媒、省级媒体、市级媒体、区级媒体及镇街媒体 6 个级别的主体的网站、App 和移动端。市级以上媒体（报、网、端），是以获得国家互联网信息办公室颁发的《互联网新闻信息服务许可证》为准入条件，佛山网络正能量指数系统包括市、省及央级信源监测数量为 253 个，没有获得许可证的（除本地部分新媒体和自媒体外），不在监测范围内。

目前，佛山网络正能量指数对 71 家市级单位、202 个区级单位及 32 个镇（街）进行定向监测。

2.2.3　*网络媒体权重因栏目不同亦有不同*

根据计算规则，媒体的权重以首页、滚动及推送为对应权重，若内容出现在二级或三级栏目，则相应向下调整到省或市级别。在统计规则里，学习强国列为重点央媒，但是，如果发布内容是入驻号或地方栏目发布，权重则按市级计算。

3. 人工校正和机器学习

“佛山正能量指数”发布已经差不多两年时间，从“试用期”（试运行）到“转正”，花了半年多的时间，期间一直在征询各个单位部门的意见，系统参数也根据各方意见在不断修正、查漏补缺，力求做到公平、公正。

关键词规则的设计是整个系统中最为重要的环节，这关乎于内容获取的准确性、完整性和权重等方面。因此，在关键词规则配置完成后，再通过人工修正、机器学习的方式来进行优化。

第一阶段：固定模型，通过人工方式，选择相应的关键词，让系统根据固定的算法自动获取内容，同时积累词汇库。这一阶段以大范围获取为主，但是准确率会偏低。

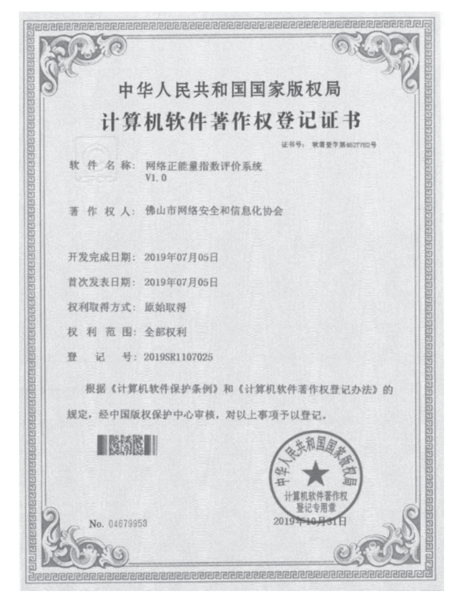
第二阶段：由系统自动判别出“确选”和“拟选”内容，通过人工对拟选内容进行筛选，再返回给系统，系统根据评估来自动优化输出结果。只要通过较长的时间和大量数据训练后，输出的精确度则会越来越高。

从一年来全社会的热点、本市的热门事件等，与相应指数反映出来的状态，也是高度吻合的（见图 5）。



图 5

经过数轮的修正和机器学习，试运行结束时，指数系统抓取的内容命中率可以达到 90% 以上，也就是说，经过系统统计出来的各级媒体的报道中，属于正能量的报道，占 90% 以上。“佛山网络正能量指数”也逐渐受到了更广泛的关注，2019 年 10 月底，“网络正能量指数评价系统”拿到了国家版权局的软件著作权证书（见图 6）。媒



中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第4627782号

软 件 名 称：网络正能量指数评价系统 V1.0

著 作 权 人：佛山市网络安全和信息化协会

开发完成日期：2019年07月05日

首次发表日期：2019年07月05日

权利取得方式：原始取得

权 利 范 围：全部权利

登 记 号：2019SR1107025

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。

中华人民共和国国家版权局 计算机软件著作权登记专用章

No. 04679953

2019年10月31日

图 6

参考文献

[1] 人民网．“党建网微平台”微信公众号．习近平谈建设网络强国 [EB/OL].(2019-03-24)http：//politics.people.com.cn/GB/n1/2019/0324/c1001-30991900.html.

[2] 佛山日报社官方帐号.2020 年度佛山网络正能量指数发布 [EB/OL].（2021-01-20）http：//epaper.fsonline.com.cn/fsrb/html/2021-01/20/content_35955_185146.htm.

[3] 李宁．浅谈大数据对新闻传播的影响 [J]. 中国传媒科技 2020（12）：20-22 .

作者简介：范春柏（1974-），男，广东佛山，佛山市网络安全和信息化协会副会长、秘书长，佛山市政协委员，佛山市新的社会阶层人士联合会副会长，记者，研究方向：网络信息安全。

（责任编辑：张晓婧）